

REMARKS

This amendment responds to the Final Office Action mailed August 24, 2006. In the Final Office Action the Examiner:

- rejected claims 12-17, 40-48 and 50-55 under 35 U.S.C. 102(e) as anticipated by Meyerzon et al. (US 6,547,829);
- rejected claims 18-20, 37-39 and 56-58 under 35 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (US 6,547,829) in view of Rujan et al. (US 6,976,207); and
- rejected claim 49 under 37 U.S.C. 103(a) as being unpatentable over Meyerzon et al. (US 6,547,829) in view of Lambert et al. (US Pub. No. 2002/0038350).

After entry of this amendment, the pending claims are: claims 12-20, 37-40 and 42-58.

Claim Rejections - 35 U.S.C. § 102

The Examiner has rejected claims 12-17, 40-48 and 50-55 under 35 U.S.C. 102(e) as being anticipated by *Meyerzon*. The Applicants respectfully disagree and traverse.

Meyerzon teaches a “first URL wins” methodology for resolving duplicate web pages. That is, the first time a specific document is encountered, it is filtered and indexed with the URL where it was found. See steps S21 – S25 in Figure 3 and Column 9, lines 32 – 50. If the same document is encountered again at another URL, this new URL is noted in history, but is not indexed. See steps S21 and S26 and Column 9, lines 32 – 40.

Applicants, however, teach a method where the URL that “wins” depends on the rank or score of the URL together with a hysteresis test. That is, when a crawled document has been encountered previously, the new copy of the document may become the “canonical” representative if it has a better rank and the rank is sufficiently better to justify a switch. When there is a switch to a new representative document, the new document is indexed. Applicants have amended the claims to clarify this point.

The amendments are supported by at least paragraphs [0048] (“This representative page is called the canonical page of its equivalence class...”) and [0069] (“each time a canonical page is replaced, the new canonical page is indexed”).

Meyerzon therefore does not teach

determining a representative document for the newly crawled document and the identified set of documents; and
indexing the representative document if the identified set of documents is empty or the representative document is not the original representative document.

For at least this reason claim 12 is not anticipated by *Meyerzon*. Further, claims 13-17, 40, 42-58, and 50-55 all have these limitations, so all of these claims are patentable over *Meyerzon*.

Claim Rejections - 35 U.S.C. § 103

The Examiner has rejected claims 18-20, 37-39 and 56-58 under 35 U.S.C. § 103(a) as unpatentable over *Meyerzon* in view of *Rujan*, and has rejected claim 49 under 35 U.S.C. § 103(a) as unpatentable over *Meyerzon* in view of *Lambert*.

The same claim limitations addressed above for the rejections under 35 U.S.C. § 102 appear in all of the claims here as well, and *Meyerzon* does not support these limitations. Further, neither *Rujan* nor *Lambert* fills in this missing element. These references address only document classification and “enhanced web page delivery,” and not a methodology for selecting a representative document.

Therefore, claims 18-20, 37-39, and 56-58 are all patentable over *Meyerzon*, *Rujan*, and *Lambert* in any combination.

CONCLUSION

In light of the amendments to the claims, and the arguments presented above, Applicants respectfully request that the Examiner reconsider this application with a view towards allowance. The Examiner is encouraged to call the undersigned attorney at (650) 843-4000 should any issues remain unresolved.

Respectfully submitted,

Date: November 22, 2006

/ Gary S. Williams /

31,066

Gary S. Williams
MORGAN, LEWIS & BOCKIUS LLP
2 Palo Alto Square
3000 El Camino Real, Suite 700
Palo Alto, California 94306
(650) 843-4000